JVI Journals.ASM.org

Identification of a Novel Rhabdovirus in *Spodoptera frugiperda* Cell Lines

Hailun Ma, Teresa A. Galvin, Dustin R. Glasner,* Syed Shaheduzzaman, Arifa S. Khan

Laboratory of Retroviruses, Division of Viral Products, Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Bethesda, Maryland, USA

ABSTRACT

The Sf9 cell line, derived from *Spodoptera frugiperda*, is used as a cell substrate for biological products, and no viruses have been reported in this cell line after extensive testing. We used degenerate PCR assays and massively parallel sequencing (MPS) to identify a novel RNA virus belonging to the order *Mononegavirales* in Sf9 cells. Sequence analysis of the assembled virus genome showed the presence of five open reading frames (ORFs) corresponding to the genes for the N, P, M, G, and L proteins in other rhabdoviruses and an unknown ORF of 111 amino acids located between the G- and L-protein genes. BLAST searches indicated that the *S. frugiperda* rhabdovirus (Sf-rhabdovirus) was related in a limited region of the L-protein gene to Taastrup virus, a newly discovered member of the *Mononegavirales* from a leafhopper (Hemiptera), and also to plant rhabdoviruses, particularly in the genus *Cytorhabdovirus*. Phylogenetic analysis of sequences in the L-protein gene indicated that Sf-rhabdovirus is a novel virus that branched with Taastrup virus. Rhabdovirus morphology was confirmed by transmission electron microscopy of filtered supernatant samples from Sf9 cells. Infectivity studies indicated potential transient infection by Sf-rhabdovirus in other insect cell lines, but there was no evidence of entry or virus replication in human cell lines. Sf-rhabdovirus sequences were also found in the Sf21 parental cell line of Sf9 cells but not in other insect cell lines, such as BT1-TN-$_5$B$_{1-4}$ (Tn$_5$; High Five) cells and Schneider's Drosophila line 2 [D.Mel.(2); SL2] cells, indicating a species-specific infection. The results indicate that conventional methods may be complemented by state-of-the-art technologies with extensive bioinformatics analysis for identification of novel viruses.

IMPORTANCE

The *Spodoptera frugiperda* Sf9 cell line is used as a cell substrate for the development and manufacture of biological products. Extensive testing has not previously identified any viruses in this cell line. This paper reports on the identification and characterization of a novel rhabdovirus in Sf9 cells. This was accomplished through the use of next-generation sequencing platforms, *de novo* assembly tools, and extensive bioinformatics analysis. Rhabdovirus identification was further confirmed by transmission electron microscopy. Infectivity studies showed the lack of replication of Sf-rhabdovirus in human cell lines. The overall study highlights the use of a combinatorial testing approach including conventional methods and new technologies for evaluation of cell lines for unexpected viruses and use of comprehensive bioinformatics strategies for obtaining confident next-generation sequencing results.

Baculovirus-insect cell expression systems are a robust platform for recombinant protein expression and have been used for the development of various investigational biological products and the manufacture of two U.S.-licensed viral vaccines (1). Since insects are phylogenetically distant from humans, the use of insect-derived cell lines (for example, High Five cells from *Trichoplusia ni* and Sf21 and Sf9 cells from *Spodoptera frugiperda*) for the manufacture of biological products is generally considered to have fewer safety concerns related to adventitious viruses than the use of mammalian cells. A comprehensive strategy to mitigate the risk of contamination in biologicals is still applicable for product safety. This includes extensive testing of biological products, including the starting materials, such as cell lines and cell culture reagents involved in the product manufacture, using a variety of general and specific assays at various stages of production (2). The currently recommended assays can demonstrate the absence of known agents; however, the presence of novel and inapparent viruses may not be detected by these assays.

The new technologies for broad virus detection that have emerged, such as massively parallel sequencing (MPS), virus microarrays, and long-range PCR with mass spectrometry, have resulted in the discovery of novel viruses in clinical materials, environmental samples, as well as some biological products. These methods are capable of detecting unexpected and novel viruses in biological products (3) and therefore may complement the current adventitious virus testing for characterization and evaluation of novel cell substrates (4). Additionally, MPS may be used to fill a knowledge gap regarding endogenous and indigenous viruses in species where little information exists, such as insect cells. This study describes the use of conventional methods and advanced technologies with extensive bioinformatics analysis to identify an unexpected, novel rhabdovirus in Sf9 cells that is more closely

Received 16 March 2014 **Accepted** 21 March 2014

Published ahead of print 26 March 2014

Editor: K. L. Beemon

Address correspondence to Arifa S. Khan, arifa.khan@fda.hhs.gov.

* Present address: Dustin R. Glasner, University of California Berkeley, Division of Infectious Diseases and Vaccinology, School of Public Health, Berkeley, California, USA.

doi:10.1128/JVI.00780-14

TABLE 1 Sf-rhabdovirus-specific primers in the L-protein gene

Primer	Sequence[a] (5′ → 3[prime])	Nucleotide position[b]	Expected size (bp)
Mono-1	GGCAAGGCTGTTTGGATTACTGACC	8365–8389	
Mono-2	ACAGGTTTGCAGCTAAGGAGGACA	9158–9135	792
Mono-1i	ATATGAGAGCCCCAGACACACAGCC	8571–8595	
Mono-2i	ACGATGTGGTGAGAGAAACACTCCT	9071–9046	500
Mono-3	TGGCGAGGGACTGCTTACAGAAGG	10630–10653	
Mono-4	CACAGCCGGGGGTGCAATCA	11359–11340	730
Mono-5	ACAGGAGATGCGGAAGACCCCTC	12137–12159	
Mono-6	ATCTCGCAGGTGGGACAACCCC	12962–12941	826

[a] Primer pairs were chosen from the L-protein gene using the Primer BLAST program (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).
[b] The nucleotide position is indicated from the start of the Sf-rhabdovirus genome sequence (GenBank accession number KF947078).

related to plant rhabdoviruses than to invertebrate or vertebrate rhabdoviruses and phylogenetically branches with the Taastrup virus, a newly discovered member of the family of *Mononegavirales* from a leafhopper (Hemiptera).

MATERIALS AND METHODS

Cell lines. The Sf9 cell line, a clonal isolate of Sf21 cells, which were derived from pupal ovarian tissue of *S. frugiperda*, was obtained from the American Type Culture Collection (ATCC; catalogue number CRL-1711, lot no. 58078522; Manassas, VA) and grown as an adherent culture in Grace's supplemented insect medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS; heat inactivated at 56°C for 30 min; certified for insect cells; HyClone, Logan, UT), 2 mM L-glutamine, 100 U of penicillin per ml, and 100 μg of streptomycin per ml (Quality Biologicals Inc., Gaithersburg, MD).

Frozen cells that were used without culturing for direct PCR analysis included Sf9 cells (ATCC and Invitrogen), Sf21 cells (Invitrogen), High Five cells (BT1-TN-$_5$B$_{1-4}$; *Trichoplusia ni* ovarian cells; Invitrogen), and Schneider's Drosophila line 2 cells [D.Mel.(2); SL2 cells; *Drosophila melanogaster* embryo cells; ATCC] (5).

Mammalian cell lines used for infectivity studies were obtained from ATCC. Vero, A204, and Raji cells were grown as previously described (6); MRC-5 cells were grown using the same medium used for Vero cells, except that it was supplemented with 10% FBS; and A549 cells were grown in Dulbecco's modified Eagle's minimal essential medium (Invitrogen) supplemented with 10% FBS. All media were supplemented with 2 mM L-glutamine, 100 U of penicillin per ml, and 100 μg of streptomycin per ml.

High Five cells were grown in Express Five serum-free medium (SFM) (Invitrogen). SL2 cells were grown in Schneider's Drosophila medium (Invitrogen) supplemented with 10% heat-inactivated FBS (certified for insect cells).

Transmission electron microscopy (TEM). Filtered supernatant from Sf9 cells was concentrated for evaluation by electron microscopy (EM). Negative staining at SGS Vitrology (Glasgow, United Kingdom) was done according to previously published procedures (7–10). Viral particles were quantified using thin sections of the pellet produced from 10 ml sample by low-speed clarification (11,000 × *g*, 10 min), filtration (pore size, 0.45 μm), and ultracentrifugation (150,000 × *g* for 3 h). Viral quantification was done as previously described (8, 10) and was based upon counting of 1,344 particles in 10 grid squares. For negative-staining electron microscopy, the pellet from 5 ml Sf9 cells was resuspended in 25 μl, and 2 μl was used.

Negative staining at NIAID, NIH (Rocky Mountain Laboratories, Hamilton, MT), was done using 5 μl of 131×-concentrated sample (10 μl), obtained by treating the Sf9 cell supernatant with 2% paraformaldehyde (final concentration in 1.5 ml) for 1 h at room temperature prior to ultracentrifugation at 125,000 × *g* for 20 h at 4°C.

Cyro-EM was performed by NanoImaging, Inc. (San Diego, CA), using 3 μl of Sf9 cell supernatant that had been concentrated 30 times sequentially using centrifugal filters with 1,000,000- and 300,000-molecular-weight cutoffs (Pall Corporation, Port Washington, NY).

RNA extraction, reverse transcriptase (RT) PCR, and nucleotide sequencing of cloned DNAs. Concentrated (1,000×) Sf9 cell supernatant was prepared by pooling medium from cells at confluence, filtering (pore size, 0.45 μm; Corning, Corning, NY), and ultracentrifugation at 125,000 × *g* for 20 h at 4°C. RNAs were extracted from 40 μl of 1,000×-concentrated Sf9 cell supernatant or from 116 μl of 1×-concentrated supernatant using a QIAamp viral RNA minikit (Qiagen, Germantown, MD) after DNase digestion (Promega, Madison, WI) by incubation at 37°C for 60 min. Total cell RNAs were prepared from a pellet of cultured cells or from a vial of frozen cells using an RNeasy Plus minikit (Qiagen). cDNA syntheses from supernatant RNAs or cell RNAs were done using an iScript cDNA synthesis kit (Bio-Rad, Hercules, CA).

PCR amplifications were done using TaKaRa *Ex Taq* kits (TaKaRa Bio Inc., Shiga, Japan) and published retroviral degenerate primers and conditions (degenerate primers dERV [11] and MOP [12]) or newly designed virus-specific primer pairs (Table 1) and the following PCR conditions: 94°C for 3 min and 35 cycles of 94°C for 30 s, 55°C for 1 min, and 72°C for 1 min, followed by 72°C for 10 min. PCR-amplified DNA fragments were isolated and sequenced as previously described (6). Sequence analysis was done using the BLAST (http://blast.ncbi.nlm.nih.gov/) and the Pfam (version 27.0; http://pfam.sanger.ac.uk/) programs.

Sucrose gradient analysis. One milliliter of filtered Sf9 cell supernatant was added to the top of a linear sucrose gradient (10 ml; 0% to 60%), which was centrifuged at 125,000 × *g* for 16 h. Fractions (500 μl each) were removed from the top, the density of each fraction was determined using a refractometer, and 116 μl of each fraction was used for RNA extraction (described above).

Massively parallel sequencing and bioinformatics analysis. The Sf9 cell pellet (2×10^7 cells, stored at −80°C) and filtered supernatant (12 ml, collected at passage 20 [p20]) were sent to BioReliance (Rockville, MD) for sequencing using the 454 Roche titanium technology. The supernatant was concentrated and nuclease treated prior to extraction of encapsidated nucleic acid. Total RNAs were prepared from the cell pellet. The two libraries were constructed by the Roche rapid library preparation method using poly(A)$^+$ cDNA. The cDNA library was sequenced on one, full sequencing plate on a GS FLX sequencer following the Roche sequencing procedures. The total number of quality reads obtained for the cell RNA transcriptome was 1,404,845, and the average length was 392 bases. The read number from supernatant RNA was 1,175,910, and the average read length was 243 bases. The raw sequence reads were compiled by the GS FLX sequencer instrument software (*gs*RunProcessor and *gs*RunBrowser) into .fna (FASTA format) raw data files.

Read assembly and sequence analysis in our lab were done using the *de novo* assembly and BLAST (13) tools of CLC genomics workbench software, version 5.5.2 (CLC bio, Denmark). The 13,534-bp genome (GenBank accession number KF947078) was assembled on 17 June 2013; however, BLAST searches continued until submission of the manuscript. At BioReliance, BLAST searches of the sequences against the sequences in its

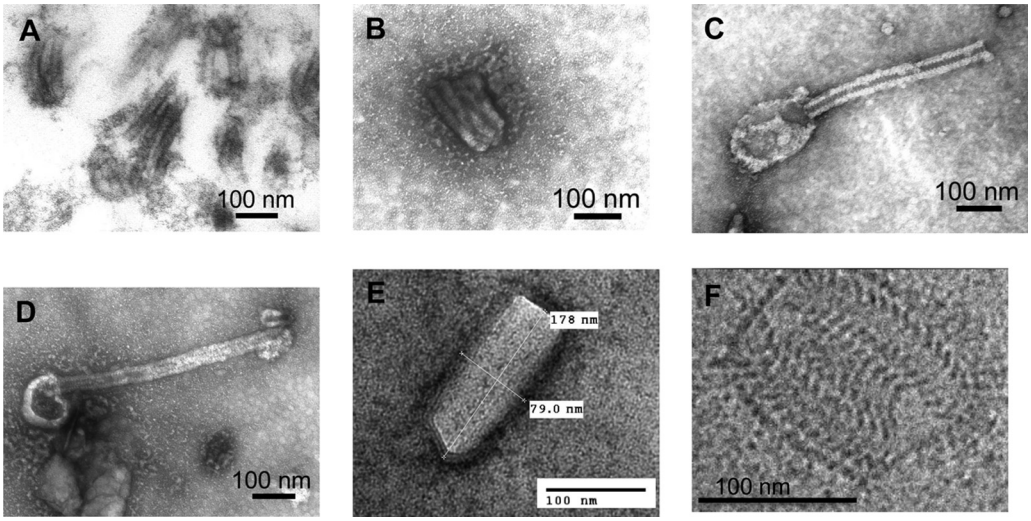

FIG 1 EM analysis of Sf9 cell supernatant. (A to D) Images of pelleted virus from Sf9 cell supernatant (SGS Vitrology). (A) A thin section stained with 2% (wt/vol) ethanolic uranyl acetate and Reynolds' lead citrate; (B to D) negative staining with 2% (wt/vol) ammonium molybdate. (E) Negative staining of pelleted virus after fixation in 2% paraformaldehyde and staining with 2% ammonium molybdate (Rocky Mountain Laboratories, NIAID, NIH). (F) Cryo-EM (NanoImaging, Inc.).

curated viral database (updated on 21 April 2012) and a nonribosomal nucleotide database (updated on 28 March 2012) were done by use of the BLASTN program.

Molecular phylogenetic analysis of L protein. Sequences for phylogenetic analysis were obtained from the database of nucleotide sequences of viruses in the NCBI nucleotide collection (nr/nt) of GenBank using a 1,500-nucleotide (nt) fragment of the *S. frugiperda* rhabdovirus (Sf-rhabdovirus; corresponding to nt 916 to 2415 in the L-protein gene), which included domain III and the adjacent upstream region that had the highest identity and the best alignment with the first 30 unique L-protein-containing sequences obtained by TBLASTX analysis. Sequences were aligned using the Muscle alignment tool with the BLOSUM substitution model in the MEGA (version 5.1) program, and then phylogenetic trees were generated using the maximum likelihood method on the basis of the Whelan and Goldman (WAG) substitution model (14). The WAG substitution model was chosen as the most appropriate for use with the test models tool in the MEGA (version 5.1) program (15). Similar trees were seen using the Jones-Taylor-Thornton (JTT) model (40).

Infectivity analysis. The inoculum for the infectivity studies was fresh supernatant collected from Sf9 cells (p20 and p24) at 95% confluence on the day of infection that had been filtered through a 0.45-μm-pore-size filter after a low-speed centrifugation at 329 × *g* for 10 min. Sf9 complete medium was used as a control. Two independent infectivity experiments were set up at 28°C and at 37°C. All target cells were split at 24 h before infection. The first infectivity experiment was set up in 25-cm^2 flasks by incubating A204, A549, and Raji cells with 2.5 ml of test sample (p20) at 37°C for 2 h before adding 2.5 ml of target cell complete medium, after which the cells were incubated for 24 h at 37°C. The second infectivity experiment was set up in 75-cm^2 flasks, in which incubating MRC-5, Vero, High Five, and SL2 cells were incubated with 7 ml of test sample (p24) at 28°C for 2 h before adding 7 ml of the target cell complete medium, after which the MRC-5 and Vero cells were incubated for 24 h at 37°C and the High Five and SL2 cells were incubated for 24 h at 28°C. Medium from all of the cultures was replaced with target cell medium after washing 3 times, and the mammalian cell lines were further incubated at 37°C and the insect cell lines were further incubated at 28°C. Cells were passaged with complete medium replacement upon reaching 95% confluence; mammalian cells were passaged every 2 to 3 days, and insect cells were passaged every 5 days. Cultures of all cells continued until day 30, except for cultures of High Five and SL2 cells, which continued until day 62.

The cultures were regularly monitored for cytopathic effect (CPE). Filtered supernatants and washed cell pellets were collected and stored at −80°C at every passage for RNA extraction and RT-PCR analysis.

Nucleotide sequence accession number. The GenBank accession number for the Sf-rhabdovirus sequence reported in this paper is KF947078.

RESULTS

Detection and identification of a novel rhabdovirus produced from Sf9 cells. Rhabdovirus-like particles were detected by TEM analysis of supernatant from Sf9 cells (Fig. 1). The number of viral particles was estimated to be 2×10^9 per ml by thin-section EM, as described in Materials and Methods. Particles seen in thin sections are shown in Fig. 1A, and negatively stained particles are shown in Fig. 1B to E. Some bacilliform viruses of about 170 to 190 nm in length and 70 nm in width were seen (Fig. 1C and D). The bacilliform viruses seemed to have a head-and-tail structure and appeared to be emerging from the envelope. Particles with the characteristic bullet-shaped morphology of rhabdoviruses were also seen (Fig. 1E). Since the heterogeneity of the structures was likely due to sample preparation for negative staining, cryo-EM was done to obtain a more accurate and detailed virus morphology (Fig. 1F). No particles were seen in the medium control.

The identity of the virus particles seen by EM was determined by RT-PCR analysis of a 1,000×-concentrated Sf9 cell supernatant with dERV and MOP degenerate primers, which were being used to investigate the RT activity produced from Sf9 cells (16, 17). Several RT-PCR-amplified products were cloned into the pGEM-T Easy vector, and nucleotide sequences were determined. Pfam program and conserved domain and database (CDD) analysis of translated protein sequences indicated that a 1,300-bp PCR-amplified DNA fragment had a high similarity (E value, 4.3e−72) with the RNA-dependent RNA polymerase (L protein) of *Mononegavirales*. A BLASTN search against the NCBI nucleotide database using the 1,300-bp fragment showed only limited similarity with Taastrup virus (GenBank accession number AY423355; 16% coverage and 66% identity; E value, −9). To obtain the full-length viral genome associated with the 1,300-bp se-

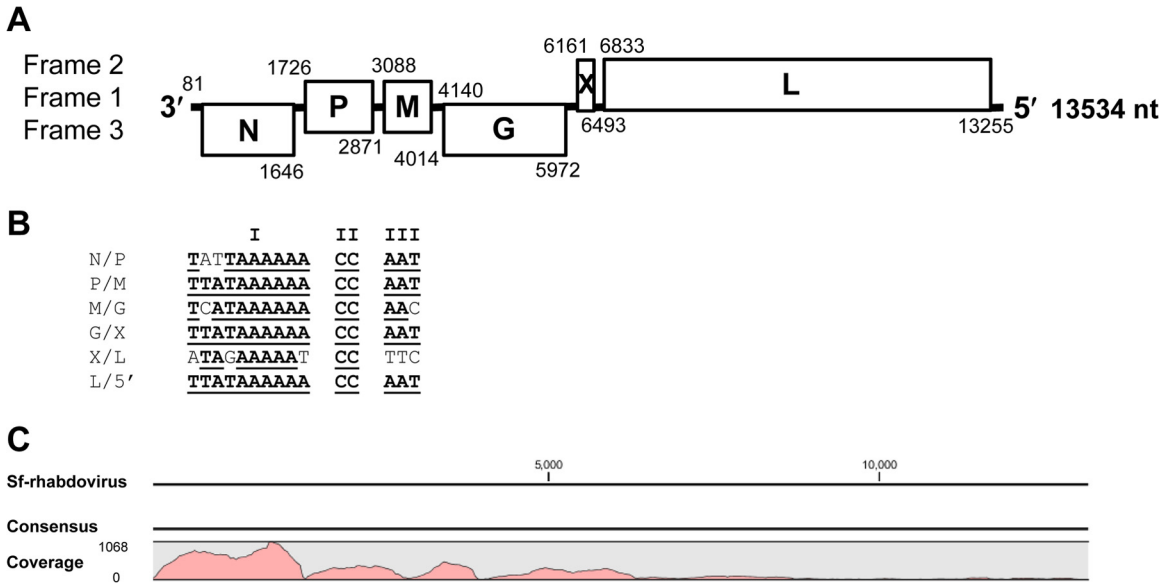

FIG 2 Genome analysis of Sf-rhabdovirus. (A) Genomic organization of Sf-rhabdovirus. The nucleotide positions of the start and end of each ORF are labeled above and below the boxes, respectively, and the reading frames and putative proteins are indicated. (B) Intergenic sequences from the plus strand corresponding to putative gene junction regions are shown. The putative polyadenylation signal, the untranscribed intergenic sequence, and the putative transcription start site are indicated I, II, and III, respectively; consensus sequences are underlined and in bold. (C) Mapping of reads obtained from MPS transcriptome data of Sf9 cells against the Sf-rhabdovirus genome sequence. The consensus sequence and coverage obtained using CLC genomics workbench software are shown.

quence, we used this fragment in the CLC genomics workbench software to interrogate the contigs generated from the reads obtained by MPS. A contig of 6,103 bp corresponding to almost the entire open reading frame (ORF) of the L protein was identified in the transcriptome of Sf9 cells. Furthermore, a contig of 5,477 bp that was different at one position (position 9624, T to A) from the 6,103-bp contig was identified in the RNAs from the Sf9 cell supernatant, but the difference did not affect the ORF. Additional virus sequences were determined by using the 6,103-bp contig as a right-hand reference sequence to obtain left-side sequences as overlapping fragments in a stepwise manner using BLASTN analysis of the contigs and reads from the transcriptomes of Sf9 cells. The assembled sequences were 13,534 bp long and contained six ORFs (each encoding proteins of >100 amino acids [aa]) and relatively conserved intergenic region sequences (Fig. 2A and B, respectively) (18, 19); five of the ORFs had the same rhabdovirus genomic organization (3′-N-P-M-G-L-5′) and were similar in length to those present in lettuce necrotic yellows virus (LNYV); additionally, there was a short ORF of 111 aa (designated X in this paper) between the genes for the G and L proteins. Since the terminal leader and trailer sequences could not be identified with confidence, the assembled Sf-rhabdovirus genome may not have been full length. Analysis of the MPS data for the RNAs from Sf9 cells indicated that 0.49% of the total reads obtained mapped with 100% coverage against the assembled Sf-rhabdovirus genome obtained from the transcriptome (Fig. 2C).

Sequence analysis of the six proteins predicted by the PSORT (version WWW.6.4) program (http://psort.hgc.jp/), recommended for use for prediction of protein localization sites in bacterial/plant/animal sequences, did not identify a nuclear localization signal with certainty.

Sucrose gradient analysis of cell-free supernatant from Sf9 cells showed that the peak of RT-PCR amplification (using primers Mono-1 and Mono-2) corresponded to fractions with densities of 1.14 to 1.17 g/ml (data not shown). These densities were similar to those of other viruses in the *Rhabdoviridae* (1.14 to 1.2 g/ml) (20, 21).

Sf-rhabdovirus sequence and phylogenetic analysis. BLASTN, TBLASTX, and BLASTX analysis of the six ORF regions of Sf-rhabdovirus showed that only sequences in the L and M regions had significant, albeit limited, similarity (E-value, <0.1) to known genes and/or proteins in the NCBI database. The BLASTN analysis of the L-protein gene (6,423 bp) showed only 4% coverage in a 295-bp region, which contained 66% identity to the L-protein gene of Taastrup virus. The TBLASTX analysis of the L-protein gene of Sf-rhabdovirus showed that the highest hit was with the translated amino acid sequences of Taastrup virus (25% coverage [1,709 bp out of 6,423 bp] and amino acid identities ranging from 31% to 47%) (Fig. 3), followed by plant cytorhabdoviruses (18) (Fig. 3). BLASTX analysis of the L protein of Sf-rhabdovirus showed 53% coverage (1,154 aa out of 2,141 aa) and 27% identity to the L protein of lettuce yellow mottle virus (LYMV); the Taastrup virus L protein was not identified since it was not in the protein database of NCBI. BLASTN and TBLASTX analysis of the M gene of Sf-rhabdovirus showed high nucleotide and translated amino acid sequence identities (99% and 94% to 98%, respectively) with 100% coverage to a *Bombyx mori* mRNA clone (GenBank accession number AK377209.1) that was obtained from an ovary-derived cell line. However, no significant similarity of the M protein to known viral proteins was found by BLASTX analysis.

Conserved domain analysis using the Pfam (version 27.0) program showed that the L protein of the novel Sf-rhabdovirus had similarity to the RNA-dependent RNA polymerase (L-protein) domain of the *Mononegavirales* (E-value, 2.1e−151). The L proteins of the *Mononegavirales* have been found to share six con-

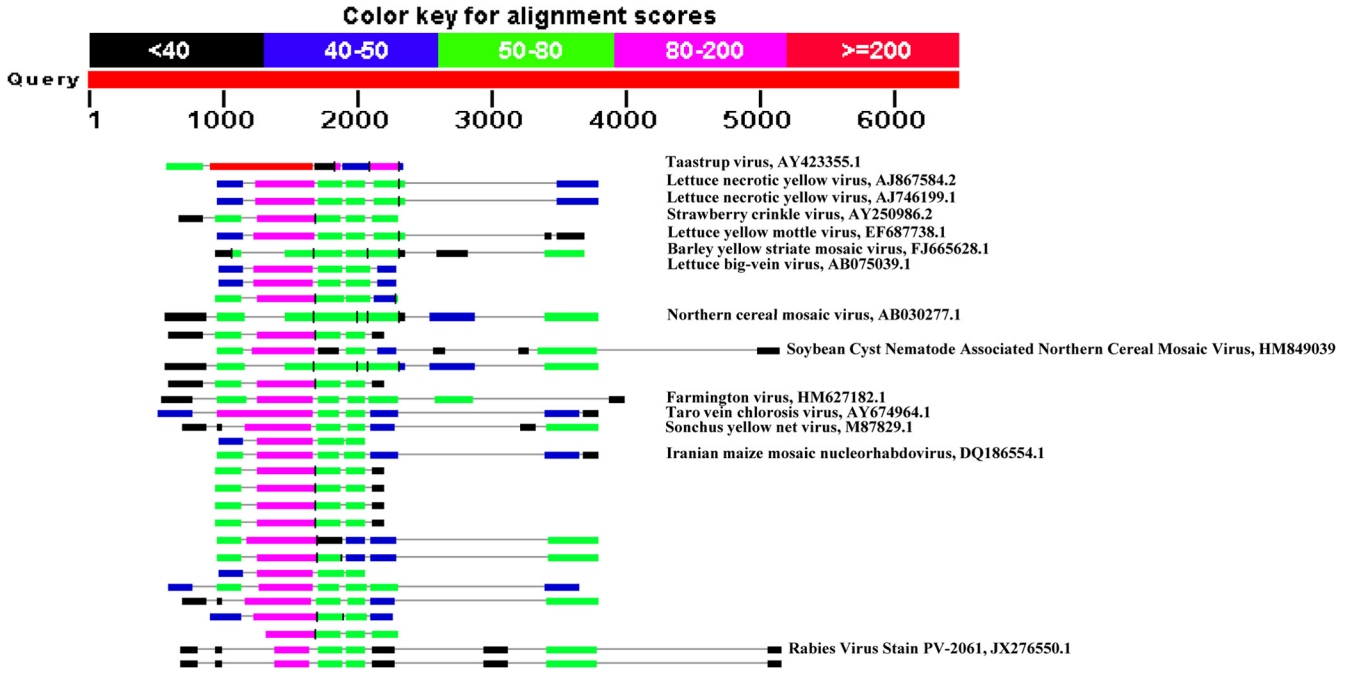

FIG 3 TBLASTX analysis of the L-protein gene of Sf-rhabdovirus. The whole sequence corresponding to the L protein was queried against the NCBI nucleotide database for viruses. Taastrup virus was the most relevant virus, with 25% coverage and 31% to 47% similarity. The color key for the alignment scores and the results of the top hits are shown. Selected viruses are labeled with their GenBank accession numbers for reference.

served domains (domains I to VI) (22, 23). Domain III is believed to play an essential role for the RNA polymerase activity (24) and contains four highly distinct and conserved motifs (22). Figure 4 shows the amino acid sequence alignment of the four motifs (motifs A to D) in the putative domain III of diverse rhabdoviruses and the Sf-rhabdovirus. These include Taastrup virus, a newly discovered member of the *Mononegavirales* (25); a rhabdovirus of *Drosophila melanogaster* grouped with the genus *Sigmavirus* (DMelSV) (26); LNYV, a plant-infecting rhabdovirus in the genus *Cytorhabdovirus* (27, 28); and rabies virus, an animal-infecting rhabdovirus in the genus *Lyssavirus* (29). The four conserved motifs of domain III are present in the Sf-rhabdovirus L protein, indicating that its potential RNA-dependent RNA polymerase function.

Due to the sequence divergence of Sf-rhabdovirus and known rhabdoviruses, phylogenetic analysis was done using a fragment

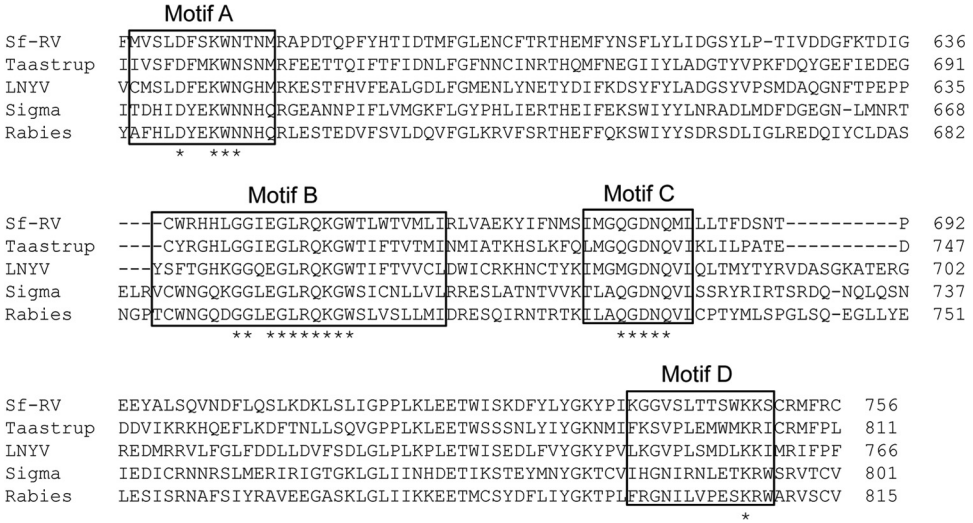

FIG 4 Alignment of domain III of the L protein of Sf-rhabdovirus (Sf-RV) with L proteins of some other viruses in the family *Rhabdoviridae*. The amino acid alignments of Sf-rhabdovirus (GenBank accession number KF947078) with Taastrup virus (Taastrup; GenBank accession number AY423355), sigma virus HAP23 from *Drosophila melanogaster* (Sigma; GenBank accession number ACU65438), lettuce necrotic yellows virus (LNYV; GenBank accession number YP_425092), and rabies virus strain PV-2061 (Rabies; GenBank accession number JX276550) are shown. Conserved motifs (motifs A to D) are boxed, and identical amino acids are indicated (*). Amino acid positions are indicated from the start of the L protein.

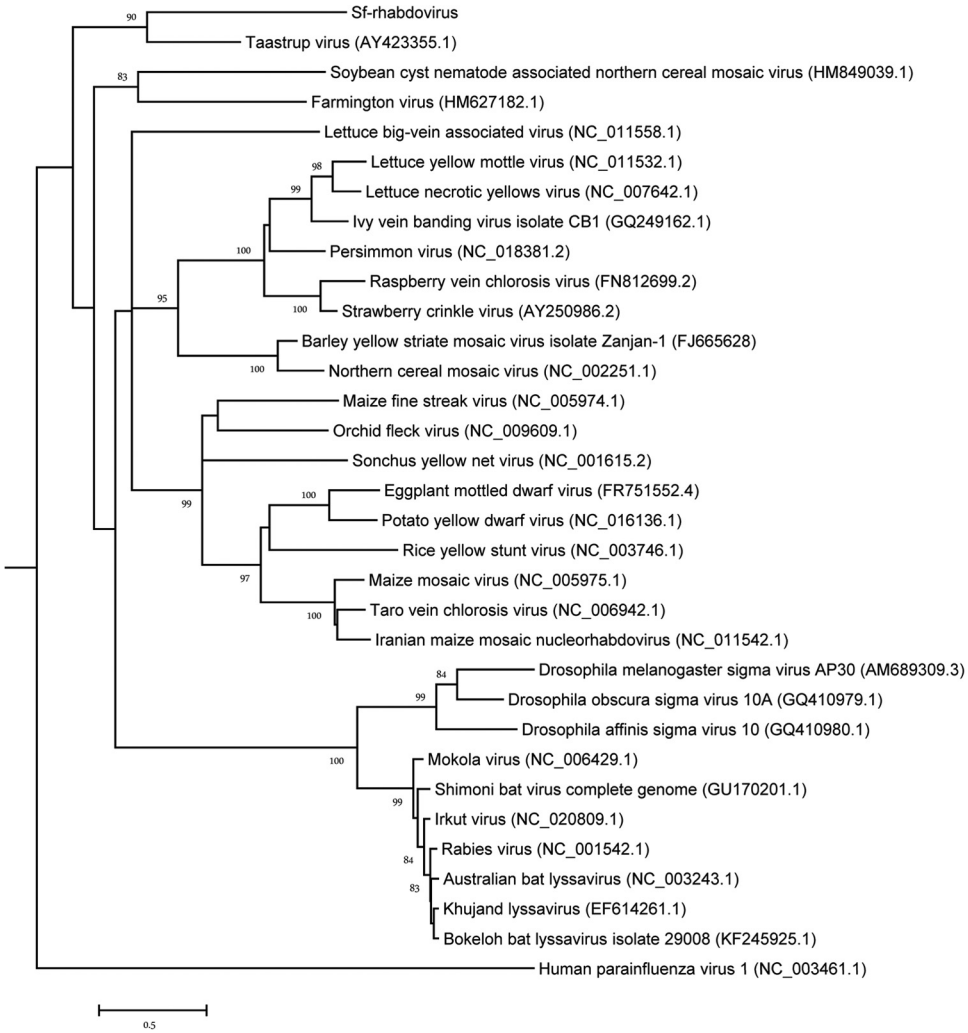

FIG 5 Molecular phylogenetic analysis of Sf-rhabdovirus. The evolutionary history was inferred by using the maximum likelihood method based on the WAG model (14). The tree with the highest log likelihood (−15,561.1192) is shown. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches (39) (values higher than 70% are indicated). The initial tree(s) for the heuristic search was obtained automatically (default settings) by applying neighbor-joining (NJ) (41) and BioNJ (42) algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories [+G, parameter = 0.9171]). The tree is drawn to scale, with branch lengths measured as the number of substitutions per site. The analysis involved 33 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 338 positions in the final data set. Evolutionary analyses were conducted in the MEGA (version 5.1) program (15). Human parainfluenza virus was used as an outlier.

that contained sequences found to have the highest identity to known viruses. These included the highly conserved domain III region of rhabdoviruses and upstream sequences that had high degrees of identity with Taastrup virus. The results showed that the Sf-rhabdovirus branched with Taastrup virus (Fig. 5). Significant phylogenetic relatedness to other rhabdoviruses was lacking due to low bootstrap values (<70). Additional phylogenetic analysis of amino acid sequences of the entire L protein of Sf-rhabdovirus with the L protein from a variety of other viruses showed results similar to those seen with the highly conserved region (data not shown).

Molecular investigations of Sf-rhabdovirus in other insect cell lines. RT-PCR analysis of RNAs extracted from Sf9 cells and from a 1,000×-concentrated supernatant using Sf-rhabdovirus-specific primers showed that L-protein gene sequences were present both in Sf9 cell RNA and in supernatant RNA (Fig. 6A). No PCR products of the expected size were found in Sf9 cell DNA (data not shown), confirming that the Sf-rhabdovirus virus is an RNA virus that does not have a stage in which it is integrated in the cell DNA.

To confirm that the Sf-rhabdovirus sequences were not introduced as a contaminant during culturing of Sf9 cells in our laboratory due to the use of animal-derived cell culture reagents, such as medium, serum, or glutamine, RT-PCR analysis was done on RNA extracted directly from original vials of Sf9 cells obtained from ATCC (designated Sf9-A) and Invitrogen (designated Sf9-I), as well as cells of the Sf21 cell line, the parental cell line of Sf9 cells. The results shown in Fig. 6B indicated that the L-protein gene sequences of Sf-rhabdovirus were similarly expressed in Sf9 cells from ATCC and Invitrogen, as well as in cells of the Sf21 cell line (lanes 2, 3, and 4, respectively). Similar analysis of insect cell lines from other species using primer sets Mono-1/Mono-2, Mono-3/

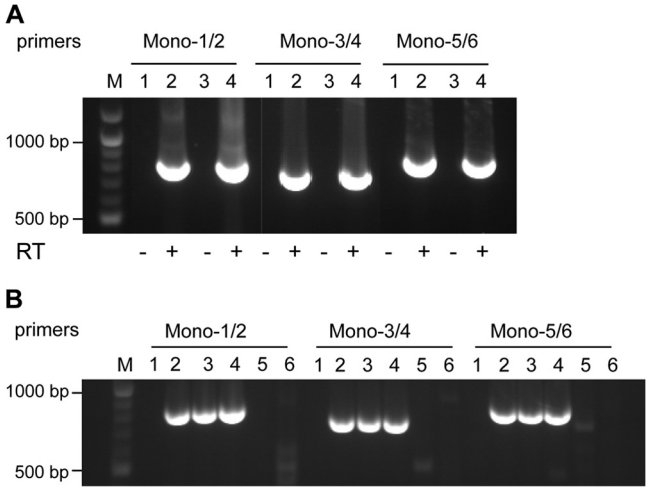

FIG 6 RT-PCR analysis of Sf-rhabdovirus-specific sequences in insect cell lines. (A) RNAs isolated from cultured Sf9 cells were analyzed by RT-PCR using the indicated primer sets in the absence (−) and presence (+) of RT. Lanes 1 and 2, cultured Sf9 cells; lanes 3 and 4, 1,000×-concentrated Sf9 cell supernatant RNA; lane M, 100-bp marker, with selected sizes indicated on the left. (B) RT-PCR analysis of RNAs from frozen insect cells. Insect cell lines were analyzed by RT-PCR using the indicated primer sets. Lanes 2 and 3, Sf9 cells from ATCC and Invitrogen, respectively; lanes 4, Sf21 cells; lanes 5, High Five cells, lanes 6, SL2 cells; lane 1, no-template negative control; lane M, 100-bp marker, with selected sizes indicated on the left.

Mono-4, and Mono-5/Mono-6 did not detect Sf-rhabdovirus sequences in High Five cells (Invitrogen) or SL2 cells (ATCC) (Fig. 6B, lanes 5 and 6, respectively). Sequence analysis of faint bands of unexpected sizes that were amplified from the RNAs of High Five and SL2 cells using Sf-rhabdovirus primers indicated a cellular origin and did not provide any evidence of Sf-rhabdovirus-related sequences.

The alignment of sequences of PCR fragments amplified in two separate reactions is shown in Fig. 7. The results showed a deletion of 6 bp in both of the Sf9 cell lines compared to the sequence of the Sf21 cell line, but the reading frame was maintained. The 6-bp deletion seemed to have resulted from a direct repeat (shown in Fig. 7), which may be indicative of Sf-rhabdovirus replication in Sf9 cells. Additionally, there were 2 single-base-pair mutations between Sf9 and Sf21 cell lines and 2 single-base-pair mutations between the ATCC Sf9 cell line and the Invitrogen Sf cell lines; however, all 4 single-base-pair mutations were silent. These results demonstrated that passage history and cell culture conditions may result in mutations in Sf-rhabdovirus.

Infectivity analysis. The replication of Sf-rhabdovirus was analyzed in human, monkey, and insect cell lines using freshly collected supernatant, without freezing and thawing, to enhance detection of an infectious virus. Additionally, incubation temperatures suitable for mammalian and insect cells were used to optimize the infection. RNAs prepared from inoculated target cells were analyzed by RT-PCR using Sf-rhabdovirus-specific nested primers. The results of the first and second PCR amplifications are shown in Fig. 8A. Sf-rhabdovirus-specific sequences were detected at day 2 in all cell lines using nested PCR (lanes 2) and at day 20 in the insect cell lines but not in the human cell lines or Vero cells (lanes 3). The persistence of Sf-rhabdovirus sequences in the insect cell lines was further evaluated by analyzing RNAs from the filtered supernatant and from cells in long-term culture. The results of the first and second amplifications are shown in Fig. 8B. The decrease in RNA expression with extended cell culture was seen in the results of the 1st PCR assays, but the results were not confirmed by quantification. Sf-rhabdovirus sequences were detected by nested PCR analysis in RNAs of High Five cells until day 62, when the cultures were terminated (lane 7), and in SL2 cells until day 34 (lane 5), whereas Sf-rhabdovirus sequences were detected in RNAs obtained from the supernatants of High Five cells at day 34 (lane 5) and from the supernatants of SL2 cells at day 6 (lane 3). These results indicate that Sf-rhabdovirus may potentially infect other insect cell lines.

DISCUSSION

Rhabdoviruses are ubiquitous in nature and have a broad host range, including invertebrates, vertebrates, and plants (30). Over 200 rhabdoviruses have been identified (31), and a majority of the plant rhabdoviruses and some vertebrate rhabdoviruses use an insect vector for transmission. However, only a few rhabdoviruses have been directly isolated from insects or insect cell lines; these belong to the order *Diptera* (such as sigma viruses in *Drosophila* and *Muscina* flies [26, 32, 33] and the *Culex tritaeniorhynchus* rhabdovirus [CTRV] in *Culex* mosquitoes [34]) and to the order *Hemiptera* (such as Taastrup virus in the leafhopper [25]). We report here the discovery of a novel rhabdovirus in the *Spodoptera*

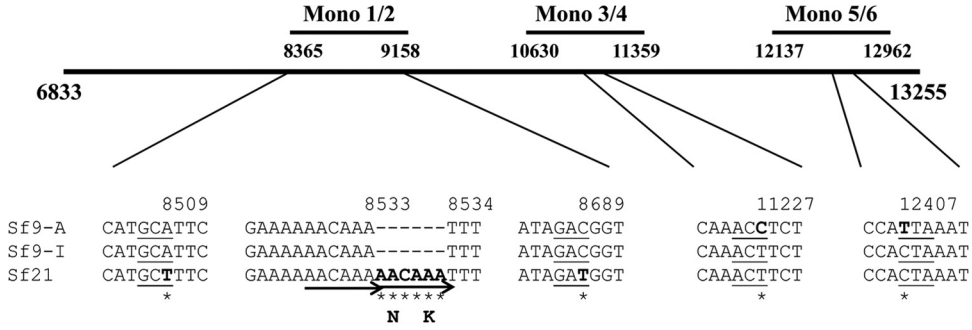

FIG 7 Comparative analysis of Sf-rhabdovirus L-protein gene sequences in Sf9 and Sf21 cells. Nucleotide sequences of Sf-rhabdovirus were obtained from 2 independent RT-PCR amplifications, using primers Mono-1/Mono-2, Mono-3/Mono-4, and Mono-5/Mono-6, from Sf9 cells from ATCC and Sf9 cells from Invitrogen (Sf9-A and Sf9-I, respectively) and cells of the Sf21 cell line. Sequences amplified from cells of each cell line in two independent PCRs were identical. The sequence alignment of the three cell lines is shown. Underlining, codons with mutations; asterisks, mutation positions; arrow, direct repeats; dashes, deleted bases. The nucleotide position on the viral genome is indicated from the 3′ terminus.

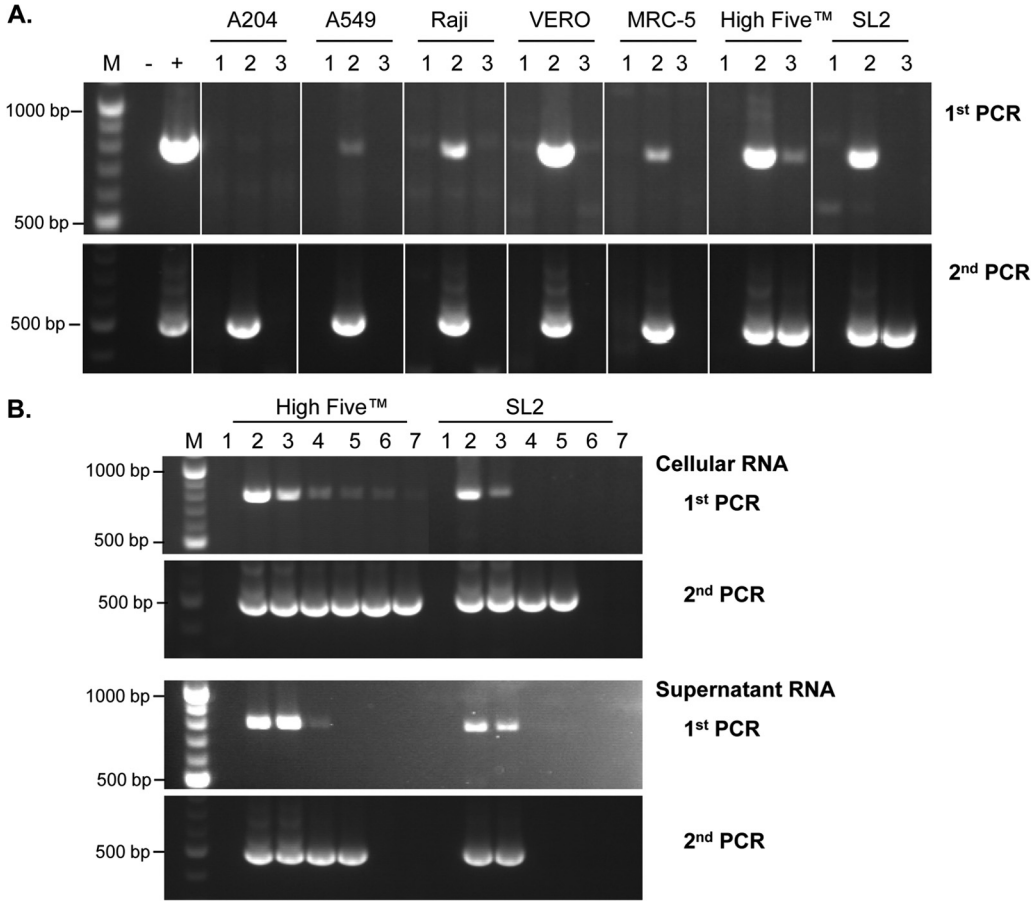

FIG 8 RT-PCR analysis of Sf-rhabdovirus sequences in inoculated target cells. RNAs prepared from cells inoculated with filtered supernatant from Sf9 cells were analyzed by RT-PCR using primers Mono-1/Mono-2 and nested primers Mono-1i/Mono-2i. (A) Cells were inoculated with Sf9 cell supernatant, and RNA was prepared on day 2 and day 20 after inoculation (lanes 2 and 3, respectively). Lanes 1, RNA prepared on day 2 from cells inoculated with complete medium as a control; lane −, no-template negative control; lane +, Sf9 cell cDNA-positive control. (B) RT-PCR analysis of cells inoculated with Sf9 cell supernatant. RNAs were prepared from cells, and filtered supernatant was obtained from inoculated High Five cells and SL2 cells on day 2, day 6, day 20, day 34, day 49, and day 62 (lanes 2 to 7, respectively). Lanes 1, RT-PCR analysis of RNAs prepared on day 2 from cells inoculated with complete medium; lane M, 100-bp marker, with selected sizes indicated on the left. Results of the first and second PCR amplifications are shown.

frugiperda Sf9 cell line, which is the first rhabdovirus found in the order *Lepidoptera*. Our initial identification of Sf-rhabdovirus was based upon the limited nucleotide sequence relatedness (4% coverage with 66% identity) of a fragment amplified from the Sf9 cells using degenerate PCR and an assembled large contig from MPS data due to highly conserved domains in the L-protein gene. The L-protein gene, which encodes the RNA-dependent RNA polymerase of *Mononegavirales*, is generally used for determining evolutionary relationships between distantly related RNA viruses (22, 23, 35). The remainder of the virus genome was determined by stepwise assembly of overlapping contigs and reads moving toward the 3′ terminus of the virus genome. Since the virus 3′ and 5′ termini were not determined, the presence of an intact virus was confirmed by EM of filtered supernatant from Sf9 cells, which showed a virus with a rhabdovirus morphology (Fig. 1).

Analysis of the Sf-rhabdovirus genome indicated that it had 5 structural genes in a linear order, 5′-N-P-M-G-L-3′, similar to other rhabdoviruses, with highly conserved intergenic regions containing putative regulatory signals, such as transcription initiation and termination signals separated by dinucleotides. The nonoverlapping ORFs with highly conserved intergenic regions identified in Sf-rhabdovirus is seen in most of the viruses within the *Rhabdoviridae* and *Paramyxoviridae* (36). BLAST searches indicated that in a limited region of the L-protein gene, Sf-rhabdovirus was highly related to Taastrup virus and was closely related to plant viruses, especially the cytorhabdoviruses (18). Phylogenetic analysis also indicated that Sf-rhabdovirus and Taastrup virus branched together, but their relatedness to other viruses could not be established because the sequence diversity resulted in low bootstrap values (<70). However, Sf-rhabdovirus was distinct from Taastrup virus, since it was not related to any known virus in any other gene except in a limited region of the L-protein gene, whereas the N protein of Taastrup virus was related to the N proteins of several members of the *Mononegavirales* (25). Furthermore, Sf-rhabdovirus was distinct from plant rhabdoviruses, since it had only one additional ORF, which was located between the G- and L-protein genes, instead of several additional ORFs, with one of those being between the P- and M-protein genes (37).

Sf-rhabdovirus was seen in Sf9 cells by EM analysis of the filtered supernatant from Sf9 cells. The large number of extracellular particles produced from the cells, which was estimated to be 2×10^9 per ml, indicated that the virus was most likely replicating in

the cells, and its persistence indicated that it was constitutively produced from the Sf9 cell line. Furthermore, the virus had established a persistent infection in Sf9 cells, since Sf-rhabdovirus sequences were also detected in the parent Sf21 cell line. Our infectivity studies indicated that Sf-rhabdovirus infection may be transient in other insect cell lines. The High Five cells, established from *Trichoplusia ni*, seemed to be more susceptible to infection than the SL2 cells established from *Drosophila melanogaster*. The susceptibility of cell lines derived from *Spodoptera* and *Trichoplusia* to Sf-rhabdovirus infection may have been because these species belong to the same Lepidoptera family (Noctuidae), whereas *Drosophila* belongs to the order Diptera. Infectivity studies indicated the absence of Sf-rhabdovirus infection using a variety of human cell lines. However, the study is based upon selection of target cell lines in the absence of knowledge about the biology of Sf-rhabdovirus. Therefore, it is prudent to demonstrate the absence of Sf-rhabdovirus in cells used for the manufacture of biological products by sensitive testing at different stages of manufacturing or incorporation of viral clearance steps in the production scheme that can be validated using relevant model viruses (2).

The identification of the Sf-rhabdovirus in Sf9 cells was unexpected, since this cell line is used in the development and manufacture of biological products, and extensive testing of the cell substrate for adventitious viruses is required (2). Our results indicate the added value of using both conventional methods and new virus detection technologies for evaluation of novel cell substrates and define a potential bioinformatics strategy for detection of novel viruses that have limited and distant relatedness to known viruses. It should be noted that the MPS data were instrumental in determining the assembled Sf-rhabdovirus genome and that other broad virus detection technologies, such as virus microarrays and broad-range PCR with mass spectrometry (Plex-ID system), were not able to detect the novel Sf-rhabdovirus, since its sequence was not related to the sequences of the primers used in this study, which were based upon known viruses or virus families (38). The bioinformatics analysis of Sf-rhabdovirus MPS data indicated the limitations of BLAST searches and available databases. Our experience with the discovery of a novel virus, Sf-rhabdovirus, underscores that only a BLASTN search of the reference genome database or a BLASTP search of the protein database would not be able to detect a sequence like Taastrup virus. The study further highlights the need for a complete, reference virus database that represents all available unique sequences for obtaining confident results from the evaluation of biological products. The results also indicate the need to further examine parameters for BLAST analysis to detect distantly related viruses. Thus, further refinement of the technologies, bioinformatics pipelines, and databases is needed to provide confidence in the detection of novel viruses.

ACKNOWLEDGMENTS

For electron microscopy, we thank Forrest H. Hoyt and Elizabeth R. Fischer at the Research Technologies Branch, Rocky Mountain Laboratories, NIAID, NIH; Yamei Gao at CBER, FDA; Euan Milne at SGS Vitrology; and Bridget Carragher and Clint Potter at NanoImaging, Inc.; and we thank Kurt Wollenberg (NIAID, NIH) for helpful discussions on sequence alignment and construction of phylogenetic trees. We thank Robin Levis and Hana Golding for critical readings of the manuscript.

The project was funded by FDA's Medical Countermeasures Initiative.

The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. The findings and conclusions in this article have not been formally disseminated by the U.S. Food and Drug Administration and should not be construed to represent any agency determination or policy.

REFERENCES

1. **U.S. Food and Drug Administration.** 2013. Vaccines, blood & biologics. U.S. Food and Drug Administration, Bethesda, MD. http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM093833.
2. **U.S. Food and Drug Administration.** 2010. Guidance for industry: characterization and qualification of cell substrates and other biological materials used in the production of viral vaccines for infectious disease indications. U.S. Food and Drug Administration, Bethesda, MD. http://www.fda.gov/downloads/biologicsbloodvaccines/guidancecomplianceregulatoryinformation/guidances/vaccines/ucm202439.pdf.
3. **Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, Delwart EL.** 2010. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J. Virol. **84:**6033–6040. http://dx.doi.org/10.1128/JVI.02690-09.
4. **U.S. Food and Drug Administration.** 2012. FDA briefing document: Vaccines and Related Biological Products Advisory Committee meeting: cell lines derived from human tumors for vaccine manufacture. U.S. Food and Drug Administration, Bethesda, MD. http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/BloodVaccinesandOtherBiologics/VaccinesandRelatedBiologicalProductsAdvisoryCommittee/UCM319573.pdf.
5. **Schneider I.** 1972. Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. **27:**353–365.
6. **Ma H, Ma Y, Ma W, Williams DK, Galvin TA, Khan AS.** 2011. Chemical induction of endogenous retrovirus particles from the Vero cell line of African green monkeys. J. Virol. **85:**6579–6588. http://dx.doi.org/10.1128/JVI.00147-11.
7. **Miller II, Mahlon F.** 1974. Particle counting of viruses, vol 4. Van Nostrand Reinhold Co, New York, NY.
8. **Milne EW.** 2003. Electron microscopy: current techniques used in product safety. BioProcessing J. **2:**65–68. http://dx.doi.org/10.12665/J22.Milne.
9. **Biel SS, Gelderblom HR.** 1999. Electron microscopy of viruses, p 111–148. *In*Cann AJ (ed), Virus culture: a practical approach. Oxford University Press, Oxford, United Kingdom.
10. **Reid GG, Milne EW, Coggins LW, Wilson NJ, Smith KT, Shepherd AJ.** 2003. Comparison of electron microscopic techniques for enumeration of endogenous retrovirus in mouse and Chinese hamster cell lines used for production of biologics. J. Virol. Methods **108:**91–96. http://dx.doi.org/10.1016/S0166-0934(02)00263-X.
11. **Donehower LA, Bohannon RC, Ford RJ, Gibbs RA.** 1990. The use of primers from highly conserved pol regions to identify uncharacterized retroviruses by the polymerase chain reaction. J. Virol. Methods **28:**33–46. http://dx.doi.org/10.1016/0166-0934(90)90085-T.
12. **Shih A, Misra R, Rush MG.** 1989. Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. J. Virol. **63:**64–75.
13. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
14. **Whelan S, Goldman N.** 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. **18:**691–699. http://dx.doi.org/10.1093/oxfordjournals.molbev.a003851.
15. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
16. **Lovatt A, Black J, Galbraith D, Doherty I, Moran MW, Shepherd AJ, Griffen A, Bailey A, Wilson N, Smith KT.** 1999. High throughput detection of retrovirus-associated reverse transcriptase using an improved fluorescent product enhanced reverse transcriptase assay and its comparison to conventional detection methods. J. Virol. Methods **82:**185–200. http://dx.doi.org/10.1016/S0166-0934(99)00111-1.
17. **Cox MM, Hollister JR.** 2009. FluBlok, a next generation influenza vaccine manufactured in insect cells. Biologicals **37:**182–189. http://dx.doi.org/10.1016/j.biologicals.2009.02.014.

18. **Redinbaugh MG, Hogenhout SA.** 2005. Plant rhabdoviruses. Curr. Top. Microbiol. Immunol. **292:**143–163. http://dx.doi.org/10.1007/3-540-27485-5_7.
19. **Conzelmann KK.** 1998. Nonsegmented negative-strand RNA viruses: genetics and manipulation of viral genomes. Annu. Rev. Genet. **32:**123–162. http://dx.doi.org/10.1146/annurev.genet.32.1.123.
20. **International Committee on Taxonomy of Viruses.** 2006. Rhabdoviridae. *In* Buchen-Osmond C (ed), ICTVdB—the universal virus database, version 3. Columbia University, New York, NY.
21. **King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed).** 2012. Virus taxonomy: classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA. http://www.ictvonline.org/.
22. **Poch O, Blumberg BM, Bougueleret L, Tordo N.** 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. **71**(Pt 5)**:**1153–1162. http://dx.doi.org/10.1099/0022-1317-71-5-1153.
23. **Poch O, Sauvaget I, Delarue M, Tordo N.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8:**3867–3874.
24. **Schnell MJ, Conzelmann KK.** 1995. Polymerase activity of in vitro mutated rabies virus L protein. Virology **214:**522–530. http://dx.doi.org/10.1006/viro.1995.0063.
25. **Bock JO, Lundsgaard T, Pedersen PA, Christensen LS.** 2004. Identification and partial characterization of Taastrup virus: a newly identified member species of the Mononegavirales. Virology **319:**49–59. http://dx.doi.org/10.1016/j.virol.2003.10.017.
26. **Longdon B, Wilfert L, Obbard DJ, Jiggins FM.** 2011. Rhabdoviruses in two species of Drosophila: vertical transmission and a recent sweep. Genetics **188:**141–150. http://dx.doi.org/10.1534/genetics.111.127696.
27. **Heim F, Lot H, Delecolle B, Bassler A, Krczal G, Wetzel T.** 2008. Complete nucleotide sequence of a putative new cytorhabdovirus infecting lettuce. Arch. Virol. **153:**81–92. http://dx.doi.org/10.1007/s00705-007-1071-5.
28. **Dietzgen RG, Callaghan B, Wetzel T, Dale JL.** 2006. Completion of the genome sequence of lettuce necrotic yellows virus, type species of the genus Cytorhabdovirus. Virus Res. **118:**16–22. http://dx.doi.org/10.1016/j.virusres.2005.10.024.
29. **Tordo N, Benmansour A, Calisher C, Dietzgen RG, Fang RX, Jackson AO, Kurath G, Nadin-Davis S, Tesh RB, Walker PJ.** 2004. Rhabdoviridae, p 623–644. *In* Fauquet CM, Mayo MA, Maniloff J, Desselberger, Ball LA (ed), Virus taxonomy. Eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, London, United Kingdom.
30. **Kuzmin IV, Novella IS, Dietzgen RG, Padhi A, Rupprecht CE.** 2009. The rhabdoviruses: biodiversity, phylogenetics, and evolution. Infect. Genet. Evol. **9:**541–553. http://dx.doi.org/10.1016/j.meegid.2009.02.005.
31. **Dietzgen RG, Calisher CH, Kurath G, Kuzman IV, Rodriguez LL, Stone DM, Tesh RB, Tordo N, Walker PJ, Wetzel T, Whitfield AE,** 2012. Rhabdoviridae, p 654–681. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
32. **Longdon B, Obbard DJ, Jiggins FM.** 2010. Sigma viruses from three species of Drosophila form a major new clade in the rhabdovirus phylogeny. Proc. Biol. Sci. **277:**35–44. http://dx.doi.org/10.1098/rspb.2009.1472.
33. **Longdon B, Wilfert L, Osei-Poku J, Cagney H, Obbard DJ, Jiggins FM.** 2011. Host-switching by a vertically transmitted rhabdovirus in Drosophila. Biol. Lett. **7:**747–750. http://dx.doi.org/10.1098/rsbl.2011.0160.
34. **Kuwata R, Isawa H, Hoshino K, Tsuda Y, Yanase T, Sasaki T, Kobayashi M, Sawabe K.** 2011. RNA splicing in a new rhabdovirus from Culex mosquitoes. J. Virol. **85:**6185–6196. http://dx.doi.org/10.1128/JVI.00040-11.
35. **Bourhy H, Cowley JA, Larrous F, Holmes EC, Walker PJ.** 2005. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. J. Gen. Virol. **86:**2849–2858. http://dx.doi.org/10.1099/vir.0.81128-0.
36. **Harcourt BH, Tamin A, Halpin K, Ksiazek TG, Rollin PE, Bellini WJ, Rota PA.** 2001. Molecular characterization of the polymerase gene and genomic termini of Nipah virus. Virology **287:**192–201. http://dx.doi.org/10.1006/viro.2001.1026.
37. **Ammar E-D, Tsai C-W, Whitfield AE, Redinbaugh MG, Hogenhout SA.** 2009. Cellular and molecular aspects of rhabdovirus interactions with insect and plant hosts. Annu. Rev. Entomol. **54:**447–468. http://dx.doi.org/10.1146/annurev.ento.54.110807.090454.
38. **Taliaferro LP, Galvin TA, Ma H, Shaheduzzaman S, William DK, Glasner DR, Khan AS.** 2014. Evaluation of the broad-range PCR-electrospray ionization mass spectrometry (PCR/ESI-MS) system and virus microarrays for virus detection. Viruses **6:**1876–1896. http://dx.doi.org/10.3390/v6051876.
39. **Felsenstein J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791. http://dx.doi.org/10.2307/2408678.
40. **Jones DT, Taylor WR, Thornton JM.** 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8:**275–282. http://dx.doi.org/10.1093/bioinformatics/8.3.275.
41. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425. http://mbe.oxfordjournals.org/content/4/4/406.long.
42. **Gascuel O.** 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. **14:**685–695. http://mbe.oxfordjournals.org/content/14/7/685.long.